



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Testing Shape Restrictions of Discrete Distributions

**Citation for published version:**

Canonne, CL, Diakonikolas, I, Gouleakis, T & Rubinfeld, R 2016, Testing Shape Restrictions of Discrete Distributions. in *Proceedings of the 33rd International Symposium on Theoretical Aspects of Computer Science (STACS 2016)*, 25, Leibniz International Proceedings in Informatics (LIPIcs), vol. 47, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, pp. 1-14, 33rd International Symposium on Theoretical Aspects of Computer Science, Orléans, France, 17/02/16.  
<https://doi.org/10.4230/LIPIcs.STACS.2016.25>

**Digital Object Identifier (DOI):**

[10.4230/LIPIcs.STACS.2016.25](https://doi.org/10.4230/LIPIcs.STACS.2016.25)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Proceedings of the 33rd International Symposium on Theoretical Aspects of Computer Science (STACS 2016)

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Testing Shape Restrictions of Discrete Distributions\*

Clément L. Canonne<sup>1</sup>, Ilias Diakonikolas<sup>2</sup>, Themis Gouleakis<sup>3</sup>, and Ronitt Rubinfeld<sup>4</sup>

- 1 Columbia University, 500 W 120th Street, New York NY, USA  
ccanonne@cs.columbia.edu
- 2 University of Edinburgh, 10 Crichton Street, Edinburgh, Scotland, UK  
ilias.d@ed.ac.uk
- 3 CSAIL, MIT, 32 Vassar Street, Cambridge MA, USA  
tgoule@mit.edu
- 4 CSAIL, MIT, 32 Vassar Street, Cambridge MA, USA; and  
the Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel  
ronitt@csail.mit.edu

---

## Abstract

We study the question of testing *structured* properties (classes) of discrete distributions. Specifically, given sample access to an arbitrary distribution  $D$  over  $[n]$  and a property  $\mathcal{P}$ , the goal is to distinguish between  $D \in \mathcal{P}$  and  $\ell_1(D, \mathcal{P}) > \varepsilon$ . We develop a general algorithm for this question, which applies to a large range of “shape-constrained” properties, including monotone, log-concave,  $t$ -modal, piecewise-polynomial, and Poisson Binomial distributions. Moreover, for all cases considered, our algorithm has near-optimal sample complexity with regard to the domain size and is computationally efficient. For most of these classes, we provide the first non-trivial tester in the literature. In addition, we also describe a generic method to prove lower bounds for this problem, and use it to show our upper bounds are nearly tight. Finally, we extend some of our techniques to tolerant testing, deriving nearly-tight upper and lower bounds for the corresponding questions.

**1998 ACM Subject Classification** F.2.2 Nonnumerical Algorithms and Problems, G.3 Probability and Statistics

**Keywords and phrases** property testing, probability distributions, statistics, lower bounds

**Digital Object Identifier** 10.4230/LIPIcs.STACS.2016.25

## 1 Introduction

Inferring information about the probability distribution that underlies a data sample is an essential question in Statistics, and one that has ramifications in every field of the natural sciences and quantitative research. In many situations, it is natural to assume that this data exhibits some simple structure because of known properties of the origin of the data, and in fact these assumptions are crucial in making the problem tractable. Such assumptions translate as constraints on the probability distribution – e.g., it is supposed to be Gaussian, or to meet a smoothness or “fat tail” condition (see e.g. [26, 24, 34]).

As a result, the problem of deciding whether a distribution possesses such a structural property has been widely investigated both in theory and practice, in the context of *shape*

---

\* The full version of this paper is available at <http://arxiv.org/abs/1507.03558>.



© Clément L. Canonne, Ilias Diakonikolas, Themis Gouleakis,  
and Ronitt Rubinfeld;  
licensed under Creative Commons License CC-BY

33rd Symposium on Theoretical Aspects of Computer Science (STACS 2016).

Editors: Nicolas Ollinger and Heribert Vollmer; Article No. 25; pp. 25:1–25:14

Leibniz International Proceedings in Informatics



Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



*restricted inference* [7, 33] and *model selection* [27]. Here, it is guaranteed or thought that the unknown distribution satisfies a shape constraint, such as having a monotone or log-concave probability density function [32, 6, 39, 19]. From a different perspective, a recent line of work in Theoretical Computer Science, originating from the papers of Batu et al. [10, 9, 22] has also been tackling similar questions in the setting of property testing (see [29, 30, 31, 12] for surveys on this field). This very active area has seen a spate of results and breakthroughs over the past decade, culminating in very efficient (both sample and time-wise) algorithms for a wide range of distribution testing problems [8, 23, 4, 18, 16, 1, 21]. In many cases, this led to a tight characterization of the number of samples required for these tasks as well as the development of new tools and techniques, drawing connections to learning and information theory [35, 36, 37].

In this paper, we focus on the following property testing problem: given a class (property) of distributions  $\mathcal{P}$  and sample access to an *arbitrary* distribution  $D$ , one must distinguish between the case that (a)  $D \in \mathcal{P}$ , versus (b)  $\|D - D'\|_1 > \varepsilon$  for all  $D' \in \mathcal{P}$  (i.e.,  $D$  is either in the class, or far from it). While many of the previous works have focused on the testing of specific properties of distributions or obtained algorithms and lower bounds on a case-by-case basis, an emerging trend in distribution testing is to design general frameworks that can be applied to *several* property testing problems [38, 21]. This direction, the testing analog of a similar movement in distribution learning [13, 15, 14, 3], aims at abstracting the minimal assumptions that are shared by a large variety of problems, and giving algorithms that can be used for any of these problems. In this work, we make significant progress in this direction by providing a unified framework for the question of testing various properties of probability distributions. More specifically, we describe a generic technique to obtain upper bounds on the sample complexity of this question, which applies to a broad range of structured classes. Our technique yields sample near-optimal and computationally efficient testers for a wide range of distribution families. Conversely, we also develop an approach – generic as well – to prove lower bounds on these sample complexities, and use it to derive tight or nearly tight bounds for many of these classes.

### Related work

Batu et al. [11] initiated the study of efficient property testers for monotonicity and obtained (nearly) matching upper and lower bounds for this problem; while [1] later considered testing the class of Poisson Binomial Distributions, and settled the sample complexity of this problem (up to the precise dependence on  $\varepsilon$ ). Indyk, Levi, and Rubinfeld [25], focusing on distributions that are piecewise constant on  $t$  intervals (“ $t$ -histograms”) described a  $\tilde{O}(\sqrt{tn}/\varepsilon^5)$ -sample algorithm for testing membership to this class. Another body of work by [8], [11], and [18] shows how assumptions on the shape of the distributions can lead to significantly more efficient algorithms. They describe such improvements in the case of identity and closeness testing as well as for entropy estimation, under monotonicity or  $k$ -modality constraints. Specifically, Batu et al. show in [11] how to obtain a  $O(\log^3 n/\varepsilon^3)$ -sample tester for closeness in this setting, in stark contrast to the  $\Omega(n^{2/3})$  general lower bound. Daskalakis et al. [18] later gave  $O(\sqrt{\log n})$  and  $O(\log^{2/3} n)$ -sample testing algorithms for testing respectively identity and closeness of monotone distributions, and obtained similar results for  $k$ -modal distributions. Finally, we briefly mention two related results, due respectively to [8] and [17]. The first one states that for the task of getting a multiplicative *estimate* of the entropy of a distribution, assuming monotonicity enables exponential savings in sample complexity –  $O(\log^6 n)$ , instead of  $\Omega(n^c)$  for the general case. The second describes how to test if an unknown  $k$ -modal distribution is in fact monotone, using only  $O(k/\varepsilon^2)$  samples. Note that the latter line of

work differs from ours in that it *presupposes* the distributions satisfy some structural property, and uses this knowledge to test something else about the distribution; while we are given *a priori* arbitrary distributions, and must *check* whether the structural property holds. Except for these two cases of monotonicity and PBDs nothing was known on testing the properties we study.<sup>1</sup>

Moreover, for the specific problems of identity and closeness testing, recent results of [21, 20] describe a general algorithm which applies to a large range of shape or structural constraints, and yields optimal identity testers for classes of distributions that satisfy them. We observe that while the question they answer can be cast as a specialized instance of membership testing, our results are incomparable to theirs, both because of the distinction above (testing *with* versus testing *for* structure) and as the structural assumptions they rely on are fundamentally different from ours.

## 1.1 Results and techniques

A natural way to tackle our membership testing problem would be to first learn the unknown distribution  $D$  *as if* it satisfied the property, before checking if the hypothesis obtained is indeed both close to the original distribution and to the property. Taking advantage of the purported structure, the first step could presumably be conducted with a small number of samples; things break down, however, in the second step. Indeed, most approximation results leading to the improved learning algorithms one would apply in the first stage only provide very weak guarantees, in the  $\ell_1$  sense. For this reason, they lack the robustness that would be required for the second part, where it becomes necessary to perform *tolerant* testing between the hypothesis and  $D$  – a task that would then entail a number of samples almost linear in  $n$ . To overcome this difficulty, we need to move away from these global  $\ell_1$  closeness results and instead work with stronger requirements, this time in  $\ell_2$  norm.

At the core of our approach is an idea of Batu et al. [11], which show that monotone distributions can be well-approximated by piecewise constant densities on a suitable partition of the domain; and leverage this fact to reduce monotonicity testing to uniformity testing on each interval of this partition. While their argument is tailored specifically for the setting of monotonicity testing, we are able to abstract the key ingredients, and generalize this idea to many classes of probability distributions. In more detail, we provide a generic testing algorithm which applies indifferently to any class of distributions which admit succinct decompositions – that is, which are well-approximated (in a strong  $\ell_2$  sense) by piecewise constant densities on a small number of intervals (we hereafter refer to this approximation property, formally defined in Definition 15, as (Succinctness); and extend the notation to apply to any *class*  $\mathcal{C}$  of distributions for which all  $D \in \mathcal{C}$  satisfy (Succinctness)). Crucially, the algorithm does not care about *how* these decompositions can be obtained: for the purpose of testing these structural properties we only need to establish their *existence*. Specific examples are given in the corollaries below. Informally, our main result is as follows:

► **Theorem 1 (Main Theorem).** *There exists an algorithm TESTSPLITTABLE which, given sampling access to an unknown distribution  $D$  over  $[n]$  and parameter  $\varepsilon \in (0, 1]$ , can distinguish with probability  $2/3$  between (a)  $D \in \mathcal{P}$  versus (b)  $\ell_1(D, \mathcal{P}) > \varepsilon$ , for any property  $\mathcal{P}$  that satisfies the above natural structural criterion (Succinctness). Moreover, for many*

<sup>1</sup> Following the communication of a preliminary version of this paper, Acharya, Daskalakis, and Kamath [2] obtained near-optimal testers for some of the classes we consider. Their work builds on ideas from [1] and, to the best of our knowledge, their techniques are orthogonal to ours.

■ **Table 1** Summary of results.

Class	Upperbound	Lowerbound
Monotone	$\tilde{O}\left(\frac{\sqrt{n}}{\varepsilon^6}\right)$ [11], $\tilde{O}\left(\frac{\sqrt{n}}{\varepsilon^{7/2}}\right)$ (Corollary 2)	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ [11], $\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
Unimodal	$\tilde{O}\left(\frac{\sqrt{n}}{\varepsilon^{7/2}}\right)$ (Corollary 2)	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
$t$ -modal	$\tilde{O}\left(\frac{\sqrt{tn}}{\varepsilon^{7/2}}\right)$ (Corollary 3)	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
Log-concave, concave, convex	$\tilde{O}\left(\frac{\sqrt{n}}{\varepsilon^{7/2}}\right)$ (Corollary 2)	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
Monotone Hazard Rate (MHR)	$\tilde{O}\left(\frac{\sqrt{n}}{\varepsilon^{7/2}}\right)$ (Corollary 2)	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
Binomial, Poisson Binomial (PBD)	$\tilde{O}\left(\frac{n^{1/4}}{\varepsilon^2} + \frac{1}{\varepsilon^6}\right)$ [1], $\tilde{O}\left(\frac{n^{1/4}}{\varepsilon^{7/2}}\right)$ (Corollary 5)	$\Omega\left(\frac{n^{1/4}}{\varepsilon^2}\right)$ ([1], Corollary 7)
$t$ -histograms	$\tilde{O}\left(\frac{\sqrt{tn}}{\varepsilon^5}\right)$ [25], $\tilde{O}\left(\frac{\sqrt{tn}}{\varepsilon^3}\right)$ (Corollary 4)	$\Omega(\sqrt{tn})$ for $t \leq \frac{1}{\varepsilon}$ [25], $\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
$t$ -piecewise degree- $d$	$\tilde{O}\left(\frac{\sqrt{t(d+1)n}}{\varepsilon^{7/2}} + \frac{t(d+1)}{\varepsilon^3}\right)$ (Corollary 4)	$\Omega\left(\frac{\sqrt{n}}{\varepsilon^2}\right)$ (Corollary 6)
$k$ -SIIRV		$\Omega(k^{1/2}n^{1/4})$ (Corollary 8)

such properties this algorithm is computationally efficient, and its sample complexity is optimal (up to logarithmic factors and the exact dependence on  $\varepsilon$ ).

We then instantiate this result to obtain “out-of-the-box” computationally efficient testers for several classes of distributions, by showing that they satisfy the premise of our theorem (the definition of these classes is given in Section 2.1):

► **Corollary 2.** *The algorithm TESTSPLITTABLE can test the classes of monotone, unimodal, log-concave, concave, convex, and monotone hazard rate (MHR) distributions, with  $\tilde{O}(\sqrt{n}/\varepsilon^{7/2})$  samples.*

► **Corollary 3.** *The algorithm TESTSPLITTABLE can test the class of  $t$ -modal distributions, with  $\tilde{O}(\sqrt{tn}/\varepsilon^{7/2})$  samples.*

► **Corollary 4.** *The algorithm TESTSPLITTABLE can test the classes of  $t$ -histograms and  $t$ -piecewise degree- $d$  distributions, with  $\tilde{O}(\sqrt{tn}/\varepsilon^3)$  and  $\tilde{O}(\sqrt{t(d+1)n}/\varepsilon^{7/2} + t(d+1)/\varepsilon^3)$  samples respectively.*

► **Corollary 5.** *The algorithm TESTSPLITTABLE can test the classes of Binomial and Poisson Binomial Distributions, with  $\tilde{O}(n^{1/4}/\varepsilon^{7/2})$  samples.*

We stress that prior to our work, no non-trivial testing bound was known for many of these classes – including  $t$ -modal, log-concave, convex, concave, MHR, and piecewise polynomials. Moreover, although three of these results are not new (the  $\tilde{O}(\sqrt{n}/\varepsilon^6)$  upper and  $\Omega(\sqrt{n}/\varepsilon^2)$  lower bounds on testing monotonicity can be found in [11], while the  $\Theta(n^{1/4})$  sample complexity of testing PBDs is pinned down<sup>2</sup> in [1] and the task of testing  $t$ -histograms is considered in [25]), the crux here is that we are able to derive them in a *unified* way, by applying the same generic algorithm to these different questions. Our result for  $t$ -histograms (Theorem 4) also improves on the previous  $\tilde{O}(\sqrt{tn}/\varepsilon^5)$ -sample tester, as long as  $t = \tilde{O}(1/\varepsilon^2)$ . As an addition to its generality, we emphasize that our framework yields a much cleaner and simpler proof of the results from [1], for both the upper and lower bounds.

<sup>2</sup> Specifically, [1] obtain an  $n^{1/4} \cdot \tilde{O}(1/\varepsilon^2) + \tilde{O}(1/\varepsilon^6)$  sample complexity, to be compared with our  $\tilde{O}(n^{1/4}/\varepsilon^{7/2}) + O(\log^4 n/\varepsilon^4)$  upper bound; as well as an  $\Omega(n^{1/4}/\varepsilon^2)$  lower bound.

## Lower bounds

As a counterpart to our testing algorithm, we give a generic framework for proving lower bounds against testing classes of distributions. In more detail, we describe how to reduce – under a mild assumption on the property  $\mathcal{C}$  – the problem of testing *membership to  $\mathcal{C}$*  (“does  $D \in \mathcal{C}$ ?”) to testing *identity to  $D^*$*  (“does  $D = D^*$ ?”), for any explicit distribution  $D^*$  in  $\mathcal{C}$ . While these two problems need not in general be related,<sup>3</sup> we show that our approach applies to a large number of natural properties and obtain nearly matching lower bounds for all of them. Moreover, this lets us derive a simple proof of the lower bound of [1] on testing the class of Poisson Binomial Distributions.

► **Corollary 6.** *Testing log-concavity, convexity, concavity, MHR, unimodality,  $t$ -modality,  $t$ -histograms, and  $t$ -piecewise degree- $d$  distributions each require  $\Omega(\sqrt{n}/\varepsilon^2)$  samples (the last three for  $t = o(\sqrt{n})$  and  $t(d+1) = o(\sqrt{n})$ , respectively), for any  $\varepsilon \geq 1/n^{O(1)}$ .*

► **Corollary 7.** *Testing the classes of Binomial and Poisson Binomial Distributions each require  $\Omega(n^{1/4}/\varepsilon^2)$  samples, for any  $\varepsilon \geq 1/n^{O(1)}$ .*

► **Corollary 8.** *There exist absolute constants  $c > 0$  and  $\varepsilon_0 > 0$  such that testing the class of  $k$ -SIIRV distributions requires  $\Omega(k^{1/2}n^{1/4})$  samples, for any  $k = o(n^c)$  and  $\varepsilon \leq \varepsilon_0$ .*

## Tolerant testing upper and lower bounds

Using our techniques, we establish the first non-trivial upper and lower bounds on tolerant testing for shape restrictions. Similarly, our upper and lower bounds are matching as a function of the domain size. More specifically, we give a generic upper bound approach (which is a simple variant of our non-tolerant testing upper bound approach). Our lower bound approach for the non-tolerant case applies to *tolerant* testing as well. In more detail, our results are as follows (see Sections 6 and 7 of the full version):

► **Corollary 9.** *Tolerant testing of log-concavity, convexity, concavity, MHR, unimodality, and  $t$ -modality can be performed with  $O\left(\frac{1}{(\varepsilon_2 - \varepsilon_1)^2} \frac{n}{\log n}\right)$  samples, for  $\varepsilon_2 \geq C\varepsilon_1$  (where  $C > 2$  is a constant).*

► **Corollary 10.** *Tolerant testing of the classes of Binomial and Poisson Binomial Distributions can be performed with  $O\left(\frac{1}{(\varepsilon_2 - \varepsilon_1)^2} \frac{\sqrt{n \log(1/\varepsilon_1)}}{\log n}\right)$  samples, for  $\varepsilon_2 \geq C\varepsilon_1$  (where  $C > 2$  is a constant).*

► **Corollary 11.** *Tolerant testing of log-concavity, convexity, concavity, MHR, unimodality, and  $t$ -modality each require  $\Omega\left(\frac{1}{(\varepsilon_2 - \varepsilon_1)} \frac{n}{\log n}\right)$  samples (the latter for  $t = o(n)$ ).*

► **Corollary 12.** *Tolerant testing of the classes of Binomial and Poisson Binomial Distributions each require  $\Omega\left(\frac{1}{(\varepsilon_2 - \varepsilon_1)} \frac{\sqrt{n}}{\log n}\right)$  samples.*

<sup>3</sup> As a simple example, consider the class  $\mathcal{C}$  of *all* distributions, for which testing membership is trivial.

<sup>3</sup> *Tolerant testing* of a property  $\mathcal{P}$  is defined as follows: given  $0 \leq \varepsilon_1 < \varepsilon_2 \leq 1$ , one must distinguish between (a)  $\ell_1(D, \mathcal{P}) \leq \varepsilon_1$  and (b)  $\ell_1(D, \mathcal{P}) \geq \varepsilon_2$ . This turns out to be, in general, a much harder task than that of “regular” testing (where we take  $\varepsilon_1 = 0$ ).

### On the scope of our results

We observe that our main theorem is likely to apply to many other classes of distributions, due to the mild structural assumptions it requires. However, we did not attempt here to be comprehensive; but rather to illustrate the generality of our results. Moreover, for all properties considered in this paper the generic upper and lower bounds we derive through our methods turn out to be optimal up to polylogarithmic factors (with regard to the support size). The reader is referred to Table 1 for a summary of our results and related work.

## 1.2 Organization of the paper

We start by giving the necessary background and definitions in Section 2, before turning to our main result, the proof of Theorem 1 (our general testing algorithm) in Section 3, and sketching our lower bound framework in Section 4. Due to space constraints, we focus in this extended abstract on proving this main theorem; and defer its applications, as well as the other sections of the paper – details on lower bounds and tolerant testing, to the full version.

## 2 Notations and preliminaries

We first restate a result of Batu et al. relating closeness to uniformity in  $\ell_2$  and  $\ell_1$  norms to “overall flatness” of the probability mass function, and which will be one of the ingredients of the proof of Theorem 1:

► **Lemma 13** ([10, 9]). *Let  $D$  be a distribution on a domain  $S$ . (a) If  $\max_{i \in S} D(i) \leq (1 + \varepsilon) \min_{i \in S} D(i)$ , then  $\|D\|_2^2 \leq (1 + \varepsilon^2)/|S|$ . (b) If  $\|D\|_2^2 \leq (1 + \varepsilon^2)/|S|$ , then  $\|D - \mathcal{U}_S\|_1 \leq \varepsilon$ .*

To check condition (b) above we shall rely on the following, which one can derive from the techniques in [21] and whose proof we defer to the full version:

► **Lemma 14** (Adapted from [21, Theorem 11]). *There exists an algorithm CHECK-SMALL- $\ell_2$  which, given parameters  $\varepsilon, \delta \in (0, 1)$  and  $c \cdot \sqrt{|I|}/\varepsilon^2 \log(1/\delta)$  independent samples from a distribution  $D$  over  $I$  (for some absolute constant  $c > 0$ ), outputs either **yes** or **no**, and satisfies the following.*

- If  $\|D - \mathcal{U}_I\|_2 > \varepsilon/\sqrt{|I|}$ , then the algorithm outputs **no** with probability at least  $1 - \delta$ ;
- If  $\|D - \mathcal{U}_I\|_2 \leq \varepsilon/2\sqrt{|I|}$ , then the algorithm outputs **yes** with probability at least  $1 - \delta$ .

## 2.1 Definitions

We give here the descriptions of the classes of distributions involved in this work (the formal definitions can be found in the full version). Recall that a distribution  $D$  over  $[n]$  is *monotone* (non-increasing) if its probability mass function (pmf) satisfies  $D(1) \geq D(2) \geq \dots D(n)$ . A natural generalization of the class  $\mathcal{M}$  of monotone distributions is the set of  $t$ -modal distributions, i.e. distributions whose pmf can go “up and down” or “down and up” up to  $t$  times.

We will also consider the classes of convex, concave, and log-concave (discrete) distributions, denoted respectively  $\mathcal{K}^+$ ,  $\mathcal{K}^-$ , and  $\mathcal{L}$  (a distribution is said to be log-concave if the logarithm of its pmf defines a concave function). It is not hard to see that convex and concave distributions are unimodal; moreover, every concave distribution is also log-concave, i.e.  $\mathcal{K}^- \subseteq \mathcal{L}$ .



Another class of interest is that of *monotone hazard rate* (MHR), which we write  $\mathcal{MHR}$ : this is the class of distributions  $D$  whose *hazard rate*  $H(i) = D(i)/\sum_{j=i}^n D(j)$  is a non-decreasing function. It is known that every log-concave distribution is both unimodal and MHR (see e.g. [5, Proposition 10]), and that monotone distributions are MHR.

Two other families of distributions have elicited significant interest in the context of density estimation, that of *histograms* (piecewise constant) and *piecewise polynomial densities*. Finally, we will also consider the two following classes, which both extend the family of Binomial distributions  $\mathcal{BIN}_n$ : on one hand, the *Poisson Binomial Distributions* ( $\mathcal{PBD}_n$ ), and on the other the *k-Sum of Independent Integer Random Variables* (*k-SIIRV*) (whose class we denote  $k\text{-SIIRV}_n$ ).

### 3 The general algorithm

In this section, we obtain our main result, restated below:

► **Theorem 1** (Main Theorem). *There exists an algorithm TESTSPLITTABLE which, given sampling access to an unknown distribution  $D$  over  $[n]$  and parameter  $\varepsilon \in (0, 1]$ , can distinguish with probability  $2/3$  between (a)  $D \in \mathcal{P}$  versus (b)  $\ell_1(D, \mathcal{P}) > \varepsilon$ , for any property  $\mathcal{P}$  that satisfies the above natural structural criterion (Succinctness). Moreover, for many such properties this algorithm is computationally efficient, and its sample complexity is optimal (up to logarithmic factors and the exact dependence on  $\varepsilon$ ).*

**Intuition.** Before diving into the proof of this theorem, we first provide a high-level description of the argument. The algorithm proceeds in 3 stages: the first, the *decomposition step*, attempts to recursively construct a partition of the domain in a small number of intervals, with a very strong guarantee. If the decomposition succeeds, then the unknown distribution  $D$  will be close (in  $\ell_1$  distance) to its “flattening” on the partition; while if it fails (too many intervals have to be created), this serves as evidence that  $D$  does not belong to the class and we can reject. The second stage, the *approximation step*, then learns this flattening of the distribution – which can be done with few samples since by construction we do not have many intervals. The last stage is purely computational, the *projection step*: where we verify that the flattening we have learned is indeed close to the class  $\mathcal{C}$ . If all three stages succeed, then by the triangle inequality it must be the case that  $D$  is close to  $\mathcal{C}$ ; and by the structural assumption on the class, if  $D \in \mathcal{C}$  then it will admit succinct enough partitions, and all three stages will go through.

Turning to the proof, we start by defining formally the “structural criterion” we shall rely on, before describing the algorithm at the heart of our result in Section 3.1. (We note that a modification of this algorithm is described in the full version, which allows us to derive Corollary 5.)

► **Definition 15** (Decompositions). Let  $\gamma > 0$  and  $L = L(\gamma, n) \geq 1$ . A class of distributions  $\mathcal{C}$  on  $[n]$  is said to be  $(\gamma, L)$ -*decomposable* if for every  $D \in \mathcal{C}$  there exists  $\ell \leq L$  and a partition  $\mathcal{I}(\gamma, D) = (I_1, \dots, I_\ell)$  of the interval  $[1, n]$  such that, for all  $j \in [\ell]$ , one of the following holds:

- (i)  $D(I_j) \leq \frac{\gamma}{L}$ ; or
- (ii)  $\max_{i \in I_j} D(i) \leq (1 + \gamma) \cdot \min_{i \in I_j} D(i)$ .

Further, if  $\mathcal{I}(\gamma, D)$  is *dyadic* (i.e., each  $I_k$  is of the form  $[j \cdot 2^i + 1, (j + 1) \cdot 2^i]$  for some integers  $i, j$ , corresponding to the leaves of a recursive bisection of  $[n]$ ), then  $\mathcal{C}$  is said to be  $(\gamma, L)$ -*splittable*.



► **Lemma 16.** *If  $\mathcal{C}$  is  $(\gamma, L)$ -decomposable, then it is  $(\gamma, O(L \log n))$ -splittable.*

### 3.1 The algorithm

Theorem 1, and with it Corollary 2 and Corollary 3 will follow from the theorem below, combined with the structural theorems (left to the full version) on the corresponding classes:

► **Theorem 17.** *Let  $\mathcal{C}$  be a class of distributions over  $[n]$  for which the following holds.*

1.  $\mathcal{C}$  is  $(\gamma, L(\gamma, n))$ -splittable;
2. *there exists a procedure  $\text{PROJECTIONDIST}_{\mathcal{C}}$  which, given as input a parameter  $\alpha \in (0, 1)$  and the explicit description of a distribution  $D$  over  $[n]$ , returns **yes** if the distance  $\ell_1(D, \mathcal{C})$  to  $\mathcal{C}$  is at most  $\alpha/10$ , and **no** if  $\ell_1(D, \mathcal{C}) \geq 9\alpha/10$  (and either **yes** or **no** otherwise).*

*Then, the algorithm  $\text{TESTSPLITTABLE}$  (Algorithm 1) is a  $O\left(\max\left(\sqrt{nL} \log n / \varepsilon^3, L / \varepsilon^2\right)\right)$ -sample tester for  $\mathcal{C}$ , for  $L = L(\varepsilon, n)$ . (Moreover, if  $\text{PROJECTIONDIST}_{\mathcal{C}}$  is computationally efficient, then so is  $\text{TESTSPLITTABLE}$ .)*

For all classes we consider, we are able to provide such computationally efficient procedures (see the full version for details).

---

#### Algorithm 1 $\text{TESTSPLITTABLE}$

---

**Require:** Domain  $I$  (interval), sample access to  $D$  over  $I$ ; subroutine  $\text{PROJECTIONDIST}_{\mathcal{C}}$

**Input:** Parameters  $\varepsilon$  and function  $L_{\mathcal{C}}(\cdot, \cdot)$ .

- 1: **SETTING UP**
  - 2:   Define  $\gamma \stackrel{\text{def}}{=} \frac{\varepsilon}{80}$ ,  $L \stackrel{\text{def}}{=} L_{\mathcal{C}}(\gamma, |I|)$ ,  $\kappa \stackrel{\text{def}}{=} \frac{\varepsilon}{160L}$ ,  $\delta \stackrel{\text{def}}{=} \frac{1}{10L}$ ; and  $c > 0$  be as in Lemma 14.
  - 3:   Set  $m \stackrel{\text{def}}{=} C \cdot \max\left(\frac{1}{\kappa}, \frac{\sqrt{L|I|}}{\varepsilon^3}\right) \cdot \log |I| = \tilde{O}\left(\frac{\sqrt{L|I|}}{\varepsilon^3} + \frac{L}{\varepsilon}\right)$    ▷  $C$  is an absolute constant.
  - 4:   Obtain a sequence  $\mathbf{s}$  of  $m$  independent samples from  $D$ .   ▷ For any  $J \subseteq I$ , let  $m_J$  be the number of samples falling in  $J$ .
  - 5:
  - 6: **DECOMPOSITION**
  - 7:   **while**  $m_I \geq \max\left(c \cdot \frac{\sqrt{|I|}}{\varepsilon^2} \log \frac{1}{\delta}, \kappa m\right)$  and at most  $L$  splits have been performed **do**
  - 8:     Run  $\text{CHECK-SMALL-}\ell_2$  (from Lemma 14) with parameters  $\frac{\varepsilon}{40}$  and  $\delta$ , using the samples of  $\mathbf{s}$  belonging to  $I$ .
  - 9:     **if**  $\text{CHECK-SMALL-}\ell_2$  outputs **no** **then**
  - 10:       Bisect  $I$ , and recurse on both halves (using the same samples).
  - 11:     **end if**
  - 12:   **end while**
  - 13:   **if** more than  $L$  splits have been performed **then**
  - 14:     **return REJECT**
  - 15:   **else**
  - 16:     Let  $\mathcal{I} \stackrel{\text{def}}{=} (I_1, \dots, I_\ell)$  be the partition of  $[n]$  from the leaves of the recursion.   ▷  $\ell \leq L$ .
  - 17:   **end if**
  - 18:
  - 19: **APPROXIMATION**
  - 20:   Learn the flattening  $\Phi(D, \mathcal{I})$  of  $D$  to  $\ell_1$  error  $\frac{\varepsilon}{20}$  (with probability  $1/10$ ), using  $O(\ell/\varepsilon^2)$  new samples. Let  $\tilde{D}$  be the resulting hypothesis.   ▷  $\tilde{D}$  is a  $\ell$ -histogram.
  - 21:
  - 22: **OFFLINE CHECK**
  - 23:   **return ACCEPT** if and only if  $\text{PROJECTIONDIST}_{\mathcal{C}}(\varepsilon, \tilde{D})$  returns **yes**.   ▷ No sample needed.
  - 24:
-

### 3.2 Proof of Theorem 17

We now give the proof of our main result (Theorem 17), first analyzing the sample complexity of Algorithm 1 before arguing its correctness. For the latter, we will need the following simple lemma from [25], restated below:

► **Fact 18** ([25, Fact 1]). *Let  $D$  be a distribution over  $[n]$ , and  $\delta \in (0, 1]$ . Given  $m \geq C \cdot \frac{\log \frac{n}{\delta}}{\eta}$  independent samples from  $D$  (for some absolute constant  $C > 0$ ), with probability at least  $1 - \delta$  we have that, for every interval  $I \subseteq [n]$ :*

- (i) *if  $D(I) \geq \frac{\eta}{4}$ , then  $\frac{D(I)}{2} \leq \frac{m_I}{m} \leq \frac{3D(I)}{2}$ ;*
- (ii) *if  $\frac{m_I}{m} \geq \frac{\eta}{2}$ , then  $D(I) > \frac{\eta}{4}$ ;*
- (iii) *if  $\frac{m_I}{m} < \frac{\eta}{2}$ , then  $D(I) < \eta$ ;*

where  $m_I = |\{j \in [m] : x_j \in I\}|$  is the number of the samples falling into  $I$ .

#### Sample complexity

The sample complexity is immediate, and comes from Steps 4 and 20. The total number of samples is

$$m + O\left(\frac{\ell}{\varepsilon^2}\right) = O\left(\frac{\sqrt{|I| \cdot L}}{\varepsilon^3} \log |I| + \frac{L}{\varepsilon} \log |I| + \frac{L}{\varepsilon^2}\right) = O\left(\frac{\sqrt{|I| \cdot L}}{\varepsilon^3} \log |I| + \frac{L}{\varepsilon^2}\right).$$

#### Correctness

Say an interval  $I$  considered during the execution of the “Decomposition” step is *heavy* if  $m_I$  is big enough on Step 7, and *light* otherwise; and let  $\mathcal{H}$  and  $\mathcal{L}$  denote the sets of heavy and light intervals respectively. By choice of  $m$  and a union bound over all  $|I|^2$  possible intervals, we can assume on one hand that with probability at least 9/10 the guarantees of Fact 18 hold simultaneously for all intervals considered. We hereafter condition on this event.

We first argue that if the algorithm does not reject in Step 13, then with probability at least 9/10 we have  $\|D - \Phi(D, \mathcal{I})\|_1 \leq \varepsilon/20$ . Indeed, we can write

$$\begin{aligned} \|D - \Phi(D, \mathcal{I})\|_1 &= \sum_{k: I_k \in \mathcal{L}} D(I_k) \cdot \|D_{I_k} - \mathcal{U}_{I_k}\|_1 + \sum_{k: I_k \in \mathcal{H}} D(I_k) \cdot \|D_{I_k} - \mathcal{U}_{I_k}\|_1 \\ &\leq 2 \sum_{k: I_k \in \mathcal{L}} D(I_k) + \sum_{k: I_k \in \mathcal{H}} D(I_k) \cdot \|D_{I_k} - \mathcal{U}_{I_k}\|_1. \end{aligned}$$

Let us bound the two terms separately.

- If  $I' \in \mathcal{H}$ , then by our choice of threshold we can apply Lemma 14 with  $\delta = \frac{1}{10L}$ ; conditioning on all of the (at most  $L$ ) events happening, which overall fails with probability at most 1/10 by a union bound, we get

$$\|D_{I'}\|_2^2 = \|D_{I'} - \mathcal{U}_{I'}\|_2^2 + \frac{1}{|I'|} \leq \left(1 + \frac{\varepsilon^2}{1600}\right) \frac{1}{|I'|}$$

as CHECK-SMALL- $\ell_2$  returned **yes**; and by Lemma 13 this implies  $\|D_{I'} - \mathcal{U}_{I'}\|_1 \leq \varepsilon/40$ .

- If  $I' \in \mathcal{L}$ , then we claim that  $D(I') \leq \max(\kappa, 2c \cdot \frac{\sqrt{|I'|}}{m\varepsilon^2} \log \frac{1}{\delta})$ . Clearly, this is true if  $D(I') \leq \kappa$ , so it only remains to show that  $D(I') \leq 2c \cdot \frac{\sqrt{|I'|}}{m\varepsilon^2} \log \frac{1}{\delta}$ . But this follows from Fact 18(i), as if we had  $D(I') > 2c \cdot \frac{\sqrt{|I'|}}{m\varepsilon^2} \log \frac{1}{\delta}$  then  $m_{I'}$  would have been big enough,

and  $I' \notin \mathcal{L}$ . Overall,

$$\sum_{I' \in \mathcal{L}} D(I') \leq \sum_{I' \in \mathcal{L}} \left( \kappa + 2c \cdot \frac{\sqrt{|I'|}}{m\epsilon^2} \log \frac{1}{\delta} \right) \leq L\kappa + 2 \sum_{I' \in \mathcal{L}} c \cdot \frac{\sqrt{|I'|}}{m\epsilon^2} \log \frac{1}{\delta} \leq \frac{\epsilon}{160} \left( 1 + \sum_{I' \in \mathcal{L}} \sqrt{\frac{|I'|}{|I|L}} \right) \leq \frac{\epsilon}{80}$$

for a sufficiently big choice of constant  $C > 0$  in the definition of  $m$ ; where we first used that  $|\mathcal{L}| \leq L$ , and then that  $\sum_{I' \in \mathcal{L}} \sqrt{\frac{|I'|}{|I|}} \leq \sqrt{L}$  by Jensen's inequality. Putting it together, this yields

$$\|D - \Phi(D, \mathcal{I})\|_1 \leq 2 \cdot \frac{\epsilon}{80} + \frac{\epsilon}{40} \sum_{I' \in \mathcal{H}} D(I_k) \leq \epsilon/40 + \epsilon/40 = \epsilon/20.$$

**Soundness.** By contrapositive, we argue that if the test returns **ACCEPT**, then (with probability at least  $2/3$ )  $D$  is  $\epsilon$ -close to  $\mathcal{C}$ . Indeed, conditioning on  $\tilde{D}$  being  $\epsilon/20$ -close to  $\Phi(D, \mathcal{I})$ , we get by the triangle inequality that

$$\begin{aligned} \|D - \mathcal{C}\|_1 &\leq \|D - \Phi(D, \mathcal{I})\|_1 + \|\Phi(D, \mathcal{I}) - \tilde{D}\|_1 + \text{dist}(\tilde{D}, \mathcal{C}) \\ &\leq \frac{\epsilon}{20} + \frac{\epsilon}{20} + \frac{9\epsilon}{10} = \epsilon. \end{aligned}$$

Overall, this happens except with probability at most  $1/10 + 1/10 + 1/10 < 1/3$ .

**Completeness.** Assume  $D \in \mathcal{C}$ . Then the choice of  $\gamma$  and  $L$  ensures the existence of a good dyadic partition  $\mathcal{I}(\gamma, D)$  in the sense of Definition 15. For any  $I$  in this partition for which (i) holds ( $D(I) \leq \frac{\gamma}{L} < \frac{\kappa}{2}$ ),  $I$  will have  $\frac{m_I}{m} < \kappa$  and be kept as a “light leaf” (this by contrapositive of Fact 18(ii)). For the other ones, (ii) holds: let  $I$  be one of these (at most  $L$ ) intervals.

- If  $m_I$  is too small on Step 7, then  $I$  is kept as “light leaf.”
- Otherwise, then by our choice of constants we can use Lemma 13 and apply Lemma 14 with  $\delta = \frac{1}{10L}$ ; conditioning on all of the (at most  $L$ ) events happening, which overall fails with probability at most  $1/10$  by a union bound, **CHECK-SMALL- $\ell_2$**  will output **yes**, as

$$\|D_I - \mathcal{U}_I\|_2^2 = \|D_I\|_2^2 - \frac{1}{|I|} \leq \left( 1 + \frac{\epsilon^2}{6400} \right) \frac{1}{|I|} - \frac{1}{|I|} = \frac{\epsilon^2}{6400|I|}$$

and  $I$  is kept as “flat leaf.”

Therefore, as  $\mathcal{I}(\gamma, D)$  is dyadic the **DECOMPOSITION** stage is guaranteed to stop within at most  $L$  splits (in the worst case, it goes on until  $\mathcal{I}(\gamma, D)$  is considered, at which point it succeeds).<sup>4</sup> Thus Step 13 passes, and the algorithm reaches the **APPROXIMATION** stage. By the foregoing discussion, this implies  $\Phi(D, \mathcal{I})$  is  $\epsilon/20$ -close to  $D$  (and hence to  $\mathcal{C}$ );  $\tilde{D}$  is then (except with probability at most  $1/10$ )  $(\frac{\epsilon}{20} + \frac{\epsilon}{20} = \frac{\epsilon}{10})$ -close to  $\mathcal{C}$ , and the algorithm returns **ACCEPT**.

<sup>4</sup> In more detail, we want to argue that if  $D$  is in the class, then a decomposition with at most  $L$  pieces is found by the algorithm. Since there is a dyadic decomposition with at most  $L$  pieces (namely,  $\mathcal{I}(\gamma, D) = (I_1, \dots, I_t)$ ), it suffices to argue that the algorithm will never split one of the  $I_j$ 's (as every single  $I_j$  will eventually be considered by the recursive binary splitting, unless the algorithm stopped recursing in this “path” before even considering  $I_j$ , which is even better). But this is the case by the above argument, which ensures each such  $I_j$  will be recognized as satisfying one of the two conditions for “good decomposition” (being either close to uniform in  $\ell_2$ , or having very little mass).

## 4 Lower bounds

We now turn to proving converses to our positive results – namely, that many of the upper bounds we obtain cannot be significantly improved upon. As in our algorithmic approach, we describe for this purpose a *generic framework* for obtaining lower bounds: roughly speaking, a “testing-by-narrowing” reduction argument, that shows that under very modest assumptions testing *membership* to a class is at least as hard as testing *identity* to any fixed distribution it contains. (Due to space constraints, we only provide here a sketch of our results – and refer the interested reader to the full version.)

### High-level idea

The motivation for our result is the observation of [11] that “monotonicity is at least as hard as uniformity.” Unfortunately, their specific argument does not generalize to other classes of distributions, making it impossible to extend it readily. The starting point of our approach is to observe that while uniformity testing is hard in general, it becomes very easy *under the promise that the distribution is monotone, or even only close to monotone*. This gives an alternate proof of the lower bound for monotonicity testing, via a different reduction: first, test if the unknown distribution is monotone; if it is, test whether it is uniform, now assuming closeness to monotone.

More generally, this idea applies to any class  $\mathcal{C}$  which (a) contains the uniform distribution, and (b) for which we have a  $o(\sqrt{n})$ -sample agnostic learner  $\mathcal{L}$ ,<sup>5</sup> as follows. Assuming we have a tester  $\mathcal{T}$  for  $\mathcal{C}$  with sample complexity  $o(\sqrt{n})$ , define a uniformity tester as below.

- test if  $D \in \mathcal{C}$  using  $\mathcal{T}$ ; if not, reject (as  $\mathcal{U} \in \mathcal{C}$ ,  $D$  cannot be uniform);
- otherwise, agnostically learn  $D$  with  $\mathcal{L}$  (since  $D$  is close to  $\mathcal{C}$ ), and obtain hypothesis  $\hat{D}$ ;
- check offline if  $\hat{D}$  is close to uniform.

By assumption,  $\mathcal{T}$  and  $\mathcal{L}$  each use  $o(\sqrt{n})$  samples, so does the whole process; but this contradicts the lower bound of [10, 28] on uniformity testing. Hence,  $\mathcal{T}$  must use  $\Omega(\sqrt{n})$  samples.

This immediately will imply Corollary 6. Moreover, this argument can be further extended to other reductions than to uniformity, which we use to derive Corollaries 7 and 8.

### The lower bound theorem

► **Theorem 19.** *Let  $\mathcal{C}$  be a class of distributions over  $[n]$  for which the following holds:*

- (i) *there exists a semi-agnostic learner  $\mathcal{L}$  for  $\mathcal{C}$ , with sample complexity  $q_L(n, \varepsilon, \delta)$  and “agnostic constant”  $c$ ;*
- (ii) *there exists a subclass  $\mathcal{C}_{\text{Hard}} \subseteq \mathcal{C}$  such that testing  $\mathcal{C}_{\text{Hard}}$  requires  $q_H(n, \varepsilon)$  samples.*

*Suppose further that  $q_L(n, \varepsilon, 1/10) = o(q_H(n, \varepsilon))$ . Then, any tester for  $\mathcal{C}$  must use  $\Omega(q_H(n, \varepsilon))$  samples.*

**Proof.** The above theorem relies on the reduction outlined above, which we rigorously detail here. Assuming  $\mathcal{C}$ ,  $\mathcal{C}_{\text{Hard}}$ ,  $\mathcal{L}$  as above (with semi-agnostic constant  $c \geq 1$ ), and a tester  $\mathcal{T}$  for

<sup>5</sup> Recall that an algorithm is said to be a *semi-agnostic learner* for a class  $\mathcal{C}$  if it satisfies the following. Given sample access to an arbitrary distribution  $D$  and parameter  $\varepsilon$ , it outputs a hypothesis  $\hat{D}$  which (with high probability) does “almost as well as it gets”:  $\|D - \hat{D}\|_1 \leq c \cdot \text{OPT}_{\mathcal{C}, D} + O(\varepsilon)$  where  $\text{OPT}_{\mathcal{C}, D} \stackrel{\text{def}}{=} \inf_{D' \in \mathcal{C}} \ell_1(D', D)$ , and  $c \geq 1$  is some absolute constant (if  $c = 1$ , the learner is said to be agnostic).

$\mathcal{C}$  with sample complexity  $q_T(n, \varepsilon)$ , we define a tester  $\mathcal{T}_{\text{Hard}}$  for  $\mathcal{C}_{\text{Hard}}$ . On input  $\varepsilon \in (0, 1]$  and given sample access to a distribution  $D$  on  $[n]$ ,  $\mathcal{T}_{\text{Hard}}$  acts as follows:

1. call  $\mathcal{T}$  with parameters  $n, \frac{\varepsilon'}{c}$  (where  $\varepsilon' \stackrel{\text{def}}{=} \frac{\varepsilon}{3}$ ) and failure probability  $1/6$ , to  $\frac{\varepsilon'}{c}$ -test if  $D \in \mathcal{C}$ . If not, reject.
2. otherwise, agnostically learn a hypothesis  $\hat{D}$  for  $D$ , with  $\mathcal{L}$  called with parameters  $n, \varepsilon'$  and failure probability  $1/6$ ;
3. check offline if  $\hat{D}$  is  $\varepsilon'$ -close to  $\mathcal{C}_{\text{Hard}}$ , accept if and only if this is the case.

We condition on both calls (to  $\mathcal{T}$  and  $\mathcal{L}$ ) to be successful, which overall happens with probability at least  $2/3$  by a union bound. The completeness is immediate: if  $D \in \mathcal{C}_{\text{Hard}} \subseteq \mathcal{C}$ ,  $\mathcal{T}$  accepts, and the hypothesis  $\hat{D}$  satisfies  $\|\hat{D} - D\|_1 \leq \varepsilon'$ . Therefore,  $\ell_1(\hat{D}, \mathcal{C}_{\text{Hard}}) \leq \varepsilon'$ , and  $\mathcal{T}_{\text{Hard}}$  accepts.

For the soundness, we proceed by contrapositive. Suppose  $\mathcal{T}_{\text{Hard}}$  accepts; it means that each step was successful. In particular,  $\ell_1(\hat{D}, \mathcal{C}) \leq \varepsilon'/c$ ; so that the hypothesis outputted by the agnostic learner satisfies  $\|\hat{D} - D\|_1 \leq c \cdot \text{OPT} + \varepsilon' \leq 2\varepsilon'$ . In turn, since the last step passed and by a triangle inequality we get, as claimed,  $\ell_1(D, \mathcal{C}_{\text{Hard}}) \leq 2\varepsilon' + \ell_1(\hat{D}, \mathcal{C}_{\text{Hard}}) \leq 3\varepsilon' = \varepsilon$ .

Observing that the overall sample complexity is  $q_T(n, \frac{\varepsilon'}{c}) + q_L(n, \varepsilon', \frac{1}{10}) = q_T(n, \frac{\varepsilon'}{c}) + o(q_H(n, \varepsilon'))$  concludes the proof.  $\blacktriangleleft$

---

## References

- 1 Jayadev Acharya and Constantinos Daskalakis. Testing Poisson Binomial Distributions. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, pages 1829–1840, 2015.
- 2 Jayadev Acharya, Constantinos Daskalakis, and Gautam C. Kamath. Optimal Testing for Properties of Distributions. In C. Cortes, N.D. Lawrence, D.D. Lee, M. Sugiyama, R. Garnett, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 3577–3598. Curran Associates, Inc., 2015. URL: <http://papers.nips.cc/paper/5839-optimal-testing-for-properties-of-distributions.pdf>.
- 3 Jayadev Acharya, Ilias Diakonikolas, Jerry Zheng Li, and Ludwig Schmidt. Sample-optimal density estimation in nearly-linear time. *CoRR*, abs/1506.00671, 2015. URL: <http://arxiv.org/abs/1506.00671>.
- 4 Noga Alon, Alexandr Andoni, Tali Kaufman, Kevin Matulef, Ronitt Rubinfeld, and Ning Xie. Testing  $k$ -wise and almost  $k$ -wise independence. In *Proceedings of the 39th ACM Symposium on Theory of Computing, STOC 2007, San Diego, California, USA, June 11-13, 2007*, pages 496–505, New York, NY, USA, 2007. doi:10.1145/1250790.1250863.
- 5 Mark Y. An. Log-concave probability distributions: theory and statistical testing. Technical report, Centre for Labour Market and Social Research, Denmark, 1996. URL: <http://EconPapers.repec.org/RePEc:fth:clmsre:96-01>.
- 6 Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic Theory*, 26(2):445–469, 2005. URL: <http://dx.doi.org/10.1007/s00199-004-0514-4>, doi:10.1007/s00199-004-0514-4.
- 7 Richard E. Barlow, Bartholomew D.J., J.M. Bremner, and H.D. Brunk. *Statistical Inference Under Order Restrictions: The Theory and Application of Isotonic Regression*. Wiley Series in Probability and Mathematical Statistics. J. Wiley, London, New York, 1972.
- 8 Tuğkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM Journal on Computing*, 35(1):132–150, 2005.
- 9 Tuğkan Batu, Eldar Fischer, Lance Fortnow, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In *42nd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2001, Las Vegas, Nevada, USA, October 14-17 2001*, pages 442–451, 2001.

- 10 Tuğkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing that distributions are close. In *41st Annual IEEE Symposium on Foundations of Computer Science, FOCS 2000, Redondo Beach, California, USA, November 12-14 2000*, pages 259–269, 2000.
- 11 Tuğkan Batu, Ravi Kumar, and Ronitt Rubinfeld. Sublinear algorithms for testing monotone and unimodal distributions. In *Proceedings of the 36th ACM Symposium on Theory of Computing, STOC 2004, Chicago, IL, USA, June 13-16, 2004*, pages 381–390, New York, NY, USA, 2004. ACM. doi:10.1145/1007352.1007414.
- 12 Clément L. Canonne. A Survey on Distribution Testing: your data is Big. But is it Blue? *Electronic Colloquium on Computational Complexity (ECCC)*, 22:63, April 2015.
- 13 Siu-on Chan, Ilias Diakonikolas, Rocco A. Servedio, and Xiaorui Sun. Learning mixtures of structured distributions over discrete domains. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1380–1394, 2013.
- 14 Siu-on Chan, Ilias Diakonikolas, Rocco A. Servedio, and Xiaorui Sun. Efficient density estimation via piecewise polynomial approximation. In *Proceedings of the 45th ACM Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 – June 03, 2014*, pages 604–613. ACM, 2014.
- 15 Siu-on Chan, Ilias Diakonikolas, Rocco A. Servedio, and Xiaorui Sun. Near-optimal density estimation in near-linear time using variable-width histograms. In *Annual Conference on Neural Information Processing Systems (NIPS)*, pages 1844–1852, 2014.
- 16 Siu-on Chan, Ilias Diakonikolas, Gregory Valiant, and Paul Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1193–1203, 2014.
- 17 Constantinos Daskalakis, Ilias Diakonikolas, and Rocco A. Servedio. Learning  $k$ -modal distributions via testing. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2012, Kyoto, Japan, January 17-19, 2012*, pages 1371–1385. Society for Industrial and Applied Mathematics (SIAM), 2012.
- 18 Constantinos Daskalakis, Ilias Diakonikolas, Rocco A. Servedio, Gregory Valiant, and Paul Valiant. Testing  $k$ -modal distributions: Optimal algorithms via reductions. In *Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2013, New Orleans, Louisiana, USA, January 6-8, 2013*, pages 1833–1852. Society for Industrial and Applied Mathematics (SIAM), 2013. URL: <http://dl.acm.org/citation.cfm?id=2627817.2627948>.
- 19 Ilias Diakonikolas. Learning structured distributions. In *Handbook of Big Data*. CRC Press, 2016.
- 20 Ilias Diakonikolas, Daniel M. Kane, and Vladimir Nikishkin. Optimal algorithms and lower bounds for testing closeness of structured distributions. In *56th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2015*, 2015.
- 21 Ilias Diakonikolas, Daniel M. Kane, and Vladimir Nikishkin. Testing Identity of Structured Distributions. In *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2015, San Diego, CA, USA, January 4-6, 2015*, 2015.
- 22 Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. Technical Report TR00-020, Electronic Colloquium on Computational Complexity (ECCC), 2000.
- 23 Sudipto Guha, Andrew McGregor, and Suresh Venkatasubramanian. Streaming and sub-linear approximation of entropy and information distances. In *Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, Miami, Florida, USA, January 22-26, 2006*, pages 733–742, Philadelphia, PA, USA, 2006. Society

- for Industrial and Applied Mathematics (SIAM). URL: <http://dl.acm.org/citation.cfm?id=1109557.1109637>.
- 24 Philip Hougaard. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*, 73:397–96, 1986.
  - 25 Piotr Indyk, Reut Levi, and Ronitt Rubinfeld. Approximating and Testing  $k$ -Histogram Distributions in Sub-linear Time. In *Proceedings of PODS*, pages 15–22, 2012.
  - 26 Benoit Mandelbrot. New methods in statistical economics. *Journal of Political Economy*, 71(5):pp. 421–440, 1963.
  - 27 Pascal Massart and Jean Picard. *Concentration inequalities and model selection*. Lecture Notes in Mathematics. Springer, 33, 2003, Saint-Flour, Cantal, 2007. URL: <http://opac.inria.fr/record=b1122538>.
  - 28 Liam Paninski. A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory*, 54(10):4750–4755, 2008.
  - 29 Dana Ron. Property Testing: A Learning Theory Perspective. *Foundations and Trends in Machine Learning*, 1(3):307–402, 2008.
  - 30 Dana Ron. Algorithmic and analysis techniques in property testing. *Foundations and Trends in Theoretical Computer Science*, 5:73–205, 2010.
  - 31 Ronitt Rubinfeld. Taming Big Probability Distributions. *XRDS*, 19(1):24–28, September 2012. doi:10.1145/2331042.2331052.
  - 32 Debasis Sengupta and Asok K. Nanda. Log-concave and concave distributions in reliability. *Naval Research Logistics (NRL)*, 46(4):419–433, 1999. doi:10.1002/(SICI)1520-6750(199906)46:4<419::AID-NAV5>3.0.CO;2-B.
  - 33 Mervyn J. Silvapulle and Pranab K. Sen. *Constrained Statistical Inference*. John Wiley & Sons, Inc., 2001. doi:10.1002/9781118165614.fmatter.
  - 34 Constantino Tsallis, Silvio V. F. Levy, André M. C. Souza, and Roger Maynard. Statistical-mechanical foundation of the ubiquity of Lévy distributions in nature. *Phys. Rev. Lett.*, 75:3589–3593, Nov 1995. doi:10.1103/PhysRevLett.75.3589.
  - 35 Gregory Valiant and Paul Valiant. A CLT and tight lower bounds for estimating entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:179, 2010.
  - 36 Gregory Valiant and Paul Valiant. Estimating the unseen: An  $n/\log n$ -sample estimator for entropy and support size, shown optimal via new clts. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 685–694, 2011.
  - 37 Gregory Valiant and Paul Valiant. An automatic inequality prover and instance optimal identity testing. In *55th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, 2014.
  - 38 Paul Valiant. Testing symmetric properties of distributions. *SIAM Journal on Computing*, 40(6):1927–1968, 2011.
  - 39 Guenther Walther. Inference and modeling with log-concave distributions. *Statistical Science*, 24(3):319–327, 2009. URL: <http://projecteuclid.org/DPubS?service=UI&version=1.0&verb=Display&handle=euclid.ss/1270041258>.